

**HHS PUBLIC ACCESS**

Author manuscript

*Commun Stat Simul Comput.* Author manuscript; available in PMC 2013 September 25.

Published in final edited form as:

*Commun Stat Simul Comput.* 1995 ; 24(3): 583–593. doi:10.1080/03610919508813260.

## A COMPARISON OF POWER APPROXIMATIONS FOR SATTERTHWAITE'S TEST

**Rachael L. DiSantostefano** and

Family Health International, P.O. Box 13950, Research Triangle Park, North Carolina 27709

**Keith E. Muller**

Department of Biostatistics, CB# 7400, University of North Carolina, Chapel Hill, North Carolina, 27599

### Abstract

When testing equality of means from two independent normal populations, many statisticians prefer heterogeneity tolerant tests. Moser, Stevens, and Watts described the noncentral density and a numerical integration algorithm for computing power. We present simple and accurate approximations for the power of the Satterthwaite test statistic. Two advantages accrue. First, the approximations substantially reduce the computational burden for tasks such as plotting power curves. Second, the approximations substantially simplify the programming and thereby make power calculations more widely available. Four methods of power approximation are evaluated for test sizes of .001, .01, .05, and .10, sample sizes of 6 and 51, variance ratios of 1 and 10, and noncentrality parameters from 0 to 50 by 1. A method based on a ratio of expected values is recommended due to its accuracy and simplicity.

### Key Words and Phrases

noncentral; t-test; Behrens-Fisher

## 1. INTRODUCTION

### 1.1 Approaches to the Behrens-Fisher Problem

Several approximations for the Behrens-Fisher problem have been proposed and evaluated on the basis of Type I and II error rates. Lee and Gurland (1975) endorsed the Welch-Aspin test based on its ability to control Type 1 error rate. Gans (1981) compared approximations to the usual t-test following a preliminary test of variance homogeneity. He concluded that the combined approach only slightly improved performance and hence recommended Welch's test.

Others have investigated the Satterthwaite statistic in detail. Zimmerman and Williams (1989) and Best and Rayner (1987) determined that the Satterthwaite statistic performs well in controlling Type I error and produces reasonable power compared to the t-test. Moser, Stevens, and Watts (1989) and Moser and Stevens (1992) compared the Satterthwaite

statistic and the usual t-test used in conjunction with a preliminary test of homogeneity. In the first paper the authors derived formulas for exact test size and power for the Always Satterthwaite (AS) test, Sometimes Satterthwaite (SS) test, and the Always T (AT) test. They also provided tables of test size and power for representative combinations of sample size, variance, and noncentrality. They concluded that the AS test produces test sizes and powers similar to the AT and SS test when the sample sizes are equal, and reasonable test sizes and powers when the sample sizes are unequal. Therefore the authors recommended always directly applying Satterthwaite's Approximate F test without any preliminary variance test.

## 1.2 Why Use An Approximation?

Moser, Stevens, and Watts' (1989) algorithm involves numerical integration of the AS density function. The computational burden becomes large in producing many power values, such as when plotting power curves. Such curves often prove very valuable in helping both statisticians and clients understand a particular power analysis. In addition, a simple approximation would avoid the need for special purpose software. Therefore an approximation with acceptable accuracy would be very useful.

## 2. DISTRIBUTIONAL PROPERTIES OF TEST STATISTICS

### 2.1 Basic Notation

Consider two sets of independent observations with

$$y_{ij} \stackrel{d}{=} N(\mu_j, \sigma_j^2) \quad (2.1.1)$$

for  $i \in \{1, \dots, N_j\}$  and  $j \in \{1, 2\}$ . Let

$$\bar{y}_j = \sum_{i=1}^{N_j} y_{ij} / N_j = \hat{\mu}_j, \quad (2.1.2)$$

$$SS_j = \sum_{i=1}^{N_j} (\bar{y}_{ij} - \bar{y}_j)^2 = (N_j - 1) \hat{\sigma}_j^2, \quad (2.1.3)$$

and

$$W_j = SS_j / [N_j(N_j - 1)]. \quad (2.1.4)$$

The null hypothesis may be stated  $H_0: \mu_1 = \mu_2$ , or equivalently,  $H_0: \mu_1 - \mu_2 = \delta = 0$ , with either a one or two-tailed alternative.

### 2.2 Independent Groups t-Test

The validity of Student's t-test requires that  $\sigma_1^2 = \sigma_2^2$ . Compute  $t^2 \stackrel{d}{=} F(1, N_1 + N_2 - 2)$ :

$$t^2 = \frac{(\hat{\mu}_1 - \hat{\mu}_2)^2}{\left(\frac{SS_1 + SS_2}{N_1 + N_2 - 2}\right) \left(\frac{1}{N_1} + \frac{1}{N_2}\right)} = \frac{\hat{\delta}^2}{\hat{\sigma} \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}. \quad (2.2.1)$$

### 2.3 Satterthwaite Test

Satterthwaite's test makes no variance homogeneity assumption. Compute

$$t_s^2 = \frac{(\hat{\mu}_1 - \hat{\mu}_2)^2}{W_1 + W_2} = \frac{\hat{\delta}^2}{W_1 + W_2} \quad (2.3.1)$$

and

$$\hat{\nu} = \frac{(W_1 + W_2)^2}{W_1^2/(N_1 - 1) + W_2^2/(N_2 - 1)}. \quad (2.3.2)$$

This test is based on assuming  $t_s^2$  is distributed as  $F(1, \nu)$ , with

$$\nu = \frac{(\sigma_1^2/N_1 + \sigma_2^2/N_2)^2}{(\sigma_1^2/N_1)^2/(N_1 - 1) + (\sigma_2^2/N_2)^2/(N_2 - 1)} \quad (2.3.3)$$

under  $H_0$ , and thus corresponds to a strategy of approximating one random variable by another. In practice  $\nu$  must be replaced (estimated) by  $\hat{\nu}$ .

### 2.4 Some Exact Distributional Results for Satterthwaite's Test

Properties of the sampling scheme allow easily proving the independence of any distinct pair among the following statistics:  $W_1$ ,  $W_2$ ,  $\hat{\mu}_1$ , and  $\hat{\mu}_2$ . It is simple to show that the numerator,  $\hat{\delta}_2 = (\hat{\mu}_1 - \hat{\mu}_2)^2$ , differs from a chi-square random variable only by a scaling constant such that

$$\hat{\delta}_2 \sim \chi^2(1, \omega), \quad (2.4.1)$$

with noncentrality parameter  $\omega = (\mu_1 - \mu_2)^2/(\sigma_1^2/N_1 + \sigma_2^2/N_2)$ . Similarly it is simple to prove that the denominator of the Satterthwaite  $t^2$  statistic,  $(W_1 + W_2)$ , is a weighted sum of independent central  $\chi^2$  statistics, with weights  $c_j = \sigma_j^2/[(N_j(N_j - 1))]$ :

$$W_1 + W_2 = \frac{\chi_{21}^2(N_1 - 1) + c_1}{\chi_{22}^2(N_2 - 1)}. \quad (2.4.2)$$

Using the exact distributions of numerator and denominator,  $t_s^2$  can be expressed as:

$$t_s^2 = \frac{\chi_{21}^2(N_1 - 1) + c_1}{\chi_{22}^2(N_2 - 1)} = \frac{\chi_{21}^2(N_1 - 1) + c_1}{\chi_{22}^2(N_2 - 1)} \cdot \frac{\chi_{21}^2(N_1 - 1) + c_1}{\chi_{22}^2(N_2 - 1)} \quad (2.4.3)$$

Moments of  $X_2$  can easily be determined using known properties of the chi-square distribution (Johnson and Kotz, 1970, p134).

$$M_1 = g(X_2) = g(W_1 + W_2) = c_1(N_1 - 1) + c_2(N_2 - 1) = (\sigma_1^2/N_1 + \sigma_2^2/N_2) \quad (2.4.4)$$

$$\begin{aligned} M_2 = \text{Var}(X_2) &= \text{Var}(W_1 + W_2) = c_1^2[2(N_1 - 1)] + c_2^2[2(N_2 - 1)] \\ &= 2[(\sigma_1^2/N_1)^2/(N_1 - 1) + (\sigma_2^2/N_2)^2/(N_2 - 1)]. \end{aligned} \quad (2.4.5)$$

## 2.5 The Satterthwaite Statistic Approximation

The Satterthwaite test corresponds to using a method of moments approximation for  $X_2 = (W_1 + W_2)$ . Define  $W = \lambda \chi^2(v)$ , so that  $g(W) = \lambda \cdot v$  and  $\text{Var}(W) = 2\lambda^2 v$ . Matching the first two moments of  $(W_1 + W_2)$  and  $W$  yields two simultaneous equations:  $M_1 = \lambda \cdot v$  and  $M_2 = 2\lambda^2 v$ . Solving the equations yields  $v = 2M_1^2/M_2$  and  $\lambda = M_2/(2M_1) = M_1/v = (\sigma_1^2/N_1 + \sigma_2^2/N_2)/v$ . Matching two moments guarantees convergence in distribution, in this case. Next define

$$\bar{\chi}_2^2(v) = \chi_1^2(1, \omega)/1 \quad \chi_2^2(v)/v. \quad (2.5.1)$$

Assuming that the distributions of  $t_s^2$  and  $t^2$  coincide defines the Satterthwaite statistic.

## 3. POWER APPROXIMATIONS

### 3.1 Notation

Let  $F_F(f, df_1, df_2, nc)$  represent the noncentral F distribution function, namely  $\Pr\{F \leq f\}$ , for a noncentral F statistic based on  $df$ , numerator  $df_1$ ,  $df_2$  denominator  $df$ , and noncentrality parameter  $nc$ . In all power approximations studied here, power for method  $m$  is computed as

$$\text{POWER}(m) = 1 - F_F(f_{\text{crit}}(m), 1, v, \omega). \quad (3.1.1)$$

The F critical value is

$$f_{\text{crit}}(m) = F_F^{-1}(1 - \alpha_t, 1, g_m \hat{v}, 0), \quad (3.1.2)$$

for  $\alpha_1$ , the target type 1 error rate and  $g_m \hat{v}$  the approximate expected value for method  $m$ .

### 3.2 Naive Approximation: $g_N \hat{v}$ .

The simplest approximation involves using the population value in lieu of the expectation:

$$g_N \hat{v} = v. \quad (3.2.1)$$

### 3.3 Univariate Repeated Measures ANOVA Approximation: $g_U \hat{v}$ .

Muller and Barton (1989, 1991) provided an approximation for the heterogeneity correction factor in the univariate approach to repeated measures. Their results can be applied to the present setting with the restriction of equal sample sizes, namely  $N_1 = N_2 = N_j$ . Define an  $N_j \times 2$  matrix,  $\tilde{Y} = (\tilde{y}_1, \tilde{y}_2)$  with covariance matrix for each row expressed as

$$\tilde{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}. \quad (3.3.1)$$

Each column of  $\tilde{Y}$  contains the responses for one group, with row pairing completely arbitrary. The usual covariance estimate is proportional to a central Wishart distribution with  $(N_j - 1)$  df and covariance given in (3.3.1) (Johnson & Kotz, 1972, p 158). Note that

$$\text{tr}(\hat{\tilde{\Sigma}}) = (\hat{\sigma}_1^2 + \hat{\sigma}_2^2), \quad (3.3.2)$$

and, with the above restriction,  $v = N_1 + N_2 - 2 = 2(N_j - 1)$ . In this setting  $\hat{v} \hat{\varepsilon}$ , with

$$\hat{\varepsilon} = \text{tr}^2(\hat{\tilde{\Sigma}}) / [2 \text{tr}(\hat{\tilde{\Sigma}})^2] = (\hat{\sigma}_1^2 + \hat{\sigma}_2^2) / 2(\hat{\sigma}_1^4 + \hat{\sigma}_2^4). \quad (3.3.3)$$

In order to conserve space, the formulae derived by Muller and Barton for the expected value of  $\hat{\varepsilon}$  are not reproduced here.

### 3.4 Ratio of Expected Values Approximation: $g_R \hat{v}$ .

Consider  $\hat{v}$  (2.3.2) as the ratio  $Q_1/Q_2$ , with  $Q_1 = (W_1 + W_2)^2$  and  $Q_2 = W_1^2/(N_1 - 1) + W_2^2/(N_2 - 1)$ . Assuming

$$g(\hat{v}) = g(Q_1)/g(Q_2) \quad (3.4.1)$$

corresponds to assuming  $Q_1$  and  $Q_2$  are independent and using a first order Taylor's series expansion for  $Q_2^{-1}$ . The exact distributional properties in Section (2.4) allows concluding that

$$g(Q_1) = \left(\frac{\sigma_1^2}{N_1}\right)^2 \frac{(N_1+1)}{(N_1-1)} + 2 \left(\frac{\sigma_1^2}{N_1}\right) \left(\frac{\sigma_2^2}{N_2}\right) + \left(\frac{\sigma_2^2}{N_2}\right) \frac{(N_2+1)}{(N_2-1)}, \quad (3.4.2)$$

and

$$g(Q_2) = \left(\frac{\sigma_1^2}{N_1}\right)^2 \frac{(N_1+1)}{(N_1-1)^2} + \left(\frac{\sigma_2^2}{N_2}\right)^2 \frac{(N_2+1)}{(N_2-1)}. \quad (3.4.3)$$

Consequently  $g_R \hat{v} = g(Q_1)/g(Q_2)$ , with  $g(Q_1)$  and  $g(Q_2)$  from (3.4.2) and (3.4.3). Appendix A contains a simple example program.

### 3.5 Taylor's Series Approximations: $g_{T(k)} \hat{v}$ .

An alternate formulation leads to a distinct Taylor's series expansion and associated approximations. Define

$$R = [SS_1 / (N_1 - 1)] / [SS_2 / (N_2 - 1)]. \quad (3.5.1)$$

Hence  $R(\sigma_2^2/\sigma_1^2)$  follows a central F distribution with  $(N_1 - 1)$  numerator and  $(N_2 - 1)$  denominator df. In turn define

$$f(R) = \hat{v} / (N_1 - 1) = (R + a_1)^2 / (R^2 + a_2)^{-1}, \quad (3.5.2)$$

in which  $a_1 = N_1/N_2$  and  $a_2 = (N_1/N_2)^2[(N_1 - 1)/(N_2 - 1)]$ . Then

$$f(R) = \sum_{k=0}^{\infty} \frac{(R - r_0)^k}{k!} f^{(k)}(r_0) = f(r_0) + (R - r_0)f^{(1)}(r_0) + \frac{(R - r_0)^2}{2!} f^{(2)}(r_0) + \dots \quad (3.5.3)$$

Choosing  $r_0 = g(R) = \mu_R$  and taking the expected value of (3.5.3) leads to a general formula for individual coefficients:

$$g(R - \mu_R)^m = g[(\sigma_1^2/\sigma_2^2)F - (\sigma_1^2/\sigma_2^2)]g(F)^m = (\sigma_1^2/\sigma_2^2)^m g[F - g(F)]^m. \quad (3.5.4)$$

Which group includes  $N_1$  observations and which contains  $N_2$  is arbitrary. Therefore group labels should be assigned such that  $N_2 \geq N_1$  because the second moment depends upon  $N_2$  and in turn accuracy will increase. Note that

$$g(\hat{v}) = (N_1 - 1)g[f(R)]. \quad (3.5.5)$$

Using 1, 2, or 3 terms yields three distinct approximations,  $g_{T(1)}\hat{v}$ ,  $g_{T(2)}\hat{v}$ , and  $g_{T(3)}\hat{v}$ . The first three moments of F (Johnson and Kotz, 1970, p190) are provided in Appendix B with the derivatives of  $f(R)$ .

## 4. NUMERICAL EVALUATION OF APPROXIMATIONS

Exact power values for the AS test were computed in SAS IMI<sup>®</sup>, using Simpson's rule for numerical integration of formulas presented in Moser, Stevens, and Watts (1989). A compressed form of the program is included in Appendix C. Exact and approximate power values were generated for  $\alpha \in \{.001, .01, .05, .10\}$ ,  $N_i \in \{6, 51\}$ ,  $i \in \{1, 2\}$ ,  $\theta = (\sigma_2^2/\sigma_1^2) \in \{1, 10\}$ , and  $\omega = (\mu_1 - \mu_2)^2 / [2(\sigma_1^2/N_1 + \sigma_2^2/N_2)] \in \{0, 1, \dots, 50\}$ . The values were chosen so that a wide range of conditions and power values were represented, including those in Moser, Stevens, and Watts (1989). The exact values were checked against results for  $\alpha = .05$  in their paper, and against results for  $\alpha = .01$  kindly supplied by the authors via personal communication. The reader should note that  $\theta$  in the numerator of equation (13) (p3967, Moser, Stevens and Watts, 1989) should be  $u$ . Also, in their Table 1, the test sizes for the AS test for  $\theta = 1$ ,  $(N_1 - 1) = 5$ , and  $(N_2 - 1) \in \{10, 50\}$  are correct only to three of the four places reported. This small deviation was verified by a simulation for  $(N_1 - 1) = 5$  and  $(N_2 - 1) = 10$  with 20 million replications: .0557 should be .0550.

Certain features of the situations evaluated should be noted. First, the size of the AS test ranges from 0.0007 to 0.0031 for  $\alpha = 0.001$ , from 0.0083 to 0.0144 for  $\alpha = 0.01$ , and from 0.0459 to 0.0550 for  $\alpha = 0.05$ . Second, differences of the form  $d = \text{approximate} - \text{exact}$  were computed for each approximation, as well as the corresponding absolute values. Third, any combination with exact power greater than .98 was eliminated from consideration to avoid bias in the order statistics. Table 1 contains descriptive statistics on the differences for  $\alpha = 0.05$ . The number of power values contributing to each tabulation (N) is reported. The maximum absolute deviation for each approximation is found in boldface. Values are correct to the four places shown.

At the  $\alpha = .05$  significance level, all approximations performed reasonably well, providing nearly two significant digits of accuracy. Considering both the absolute maximum difference as well as the sign of the difference, the Ratio approximation performed best, with a maximum deviation of  $-0.0179$  (corresponding to slight conservatism). The median difference for the Univariate Repeated and Ratio of expected values approximations equaled zero, suggesting that they under-estimated power as often as they over-estimated power. With medians above zero, the Naive and all Taylor's series approximations tended to over-estimate more often than underestimate. The Ratio and Univariate Repeated approximations achieved the smallest maxima (0.0014 and 0.0067). Maxima for the remaining approximations fell in the range of 0.0116 to 0.0353. Maximum absolute deviations ranged from a best of 0.0018, for the two and three term Taylor's series approximations, to a worst of 0.0353, for the two term Taylor's series approximation. By looking at specific combinations of  $N_1$ ,  $N_2$ ,  $\theta$ , and  $\omega$ , it is seen that the maximum absolute deviations occurred when  $N_1 = N_2 = 6$ . The Univariate Repeated achieved this maximum when  $\theta = 1$ , while the Ratio and one and two-term Taylor's series approximations achieved their maximum absolute deviation when  $\theta = 10$ . Note that the three-term Taylor's series and the Univariate Repeated approximations can only be applied to a restricted set of cases, which tends to flatter each.

A similar pattern of results held for all other values of  $\alpha$ . As  $\alpha$  decreased, accuracy decreased, and as  $\alpha$  increased, accuracy increased. All measures of performance were worst for  $\alpha = 0.001$ , with the approximations off by as much as .15. The Ratio of expected values and one term Taylor's series achieve the largest minima ( $-0.0889$  and  $-0.0761$ ). With medians below zero, these approximations as well as the univariate repeated approximation tended to under-estimate power more than they over-estimated power. On the other hand, the naive and two and three-term Taylor's series tended to over-estimate power. The smallest maximum was achieved by the Ratio approximation (0.0181). The remaining maxima ranged from 0.0524 to 0.1520.

Plotting the difference between approximate and exact power as a function of exact power revealed important information. Worst performance occurred with  $\alpha = .001$ ,  $N_1 = 6$ ,  $N_2 = 51$ ,  $\theta = 1$ . Both under-estimation and over-estimation occurred. For  $\alpha = .001$  and power  $> .70$ , the Ratio approximation differed in absolute value from true power by at most .025, for all conditions. In contrast, the three term Taylor's Series expansion over-estimated power when the exact power was above .40. The over-estimation occurred for higher powers, making it more important than the under-estimation of the Ratio approximation.

One possible source of power inaccuracy lies in approximating  $g_{v..}$ . This led us to compute the results in Table II. The fourth column simultaneously provides the value of  $v..$  and the naive approximation of  $g_{v..}$ . The column labeled “Exact” was computed via numerical integration of  $v..(u) \cdot f(u)$ , defined in equations (4) and (11) in Moser, Stevens, and Watts (1989). The table reveals that the expected value approximations may be either larger or smaller than the true value. Furthermore  $g_{v..}$  may be either larger or smaller than  $v..$ .

Exact expected values from Table II were used with equations (3.1.1) and (3.1.2) to compute approximate power. The approximation usually performed worse than the Ratio and Taylor’s series approximations. Hence the approximation of the distribution function embodied in equation (3.1.1) plays at least an equal role in power approximation.

## 5. DISCUSSION

All of the approximations share certain features. Naturally accuracy increases with sample size. Furthermore accuracy is less if fewer observations are associated with the larger variance. Less obviously, accuracy increases with  $a$ . All approximations worked well at the  $\alpha = 0.05$  and  $\alpha = 0.10$  levels, providing nearly two significant digits.

Overall, the Ratio of expected values approximation worked best. This conclusion is based not only on the absolute amount of deviation, but also on the sign and location (with respect to exact power) of any deviation. Usually one prefers modest under-estimation to modest over-estimation of power. The Ratio approximation properties best reflect the preference, as well as being somewhat easier to compute than the series approximations. Even for or  $\alpha = .001$ , nearly two digits of accuracy are available when power  $> .70$ .

Naturally the Taylor’s series could be expanded in order to speed computing (relative to the exact method). Note that the existence (and hence computability) of the  $m$ th moment of the ratio requires the largest cell size to be greater than  $(2m-1)$ . Therefore the length of a Taylor’s series is limited by the sample size.

When maximum accuracy is desired the exact methods of Moser, Stevens, and Watts should be used, especially for power  $< .70$  and  $\alpha < 0.01$ . The Ratio approximation provides a convenient and rapid approximation. In most practical settings, the accuracy should be completely acceptable.

## Acknowledgments

The authors gratefully acknowledge suggestions made by anonymous reviewers on an earlier version of this paper. Muller’s work was supported in part by NCI program project grant P01 CA47 982-04, and NIH Clinical Research Center Grant RR00046.

## APPENDIX A

Compressed Template SAS Program for Ratio of Expected Values Power Calculation



```

DATA ONE;
ALPHA=.05;
DO VAR2=1,10;
  DO VAR1=1;
    DO N1= 6,51;      NU1=N1-1;
      DO N2=6,51;      NU2=N2-1;
        POPW1=VAR1/N1;  POPW2=VAR2/N2;
        C1=POPW1/(N1-1); C2=POPW2/(N2-1);
        C1SQ=C1**2;      C2SQ=C2**2;
        EX1=C1SQ*(N1-1)*(N1+1) + 2*C1*(N1-1)*C2*(N2-1) + C2SQ*(N2-1)*(N2+1);
        EX2=C1SQ*(N1+1) + C2SQ*(N2+1) ;
        EXNUDEN=EX1/EX2;
        FCRT=FINV(1-ALPHA,1,EXNUDEN);
        NUSTAR1=(POPW1+POPW2)**2;
        NUSTAR2=POPW1*POPW1/(N1-1) + POPW2*POPW2/(N2-1);
        NUSTAR=NUSTAR1/NUSTAR2;
        DO LAMBDA=0 TO 10 BY 1;
          SATPOW =1-PROBF(FCRT,1,NUSTAR, LAMBDA);
          OUTPUT;  END;  END;  END;  END;  END;  END;
        END;
      END;
    END;
  END;
END;

```

## APPENDIX B

Derivatives of  $f(R)$  and Moments of  $F$  for Taylor's Series Approximation

$$\begin{aligned}
 f(R) &= (r+a_1)^2 (r^2+a_2)^{-1} \\
 f^{(1)}(R) &= -2r(a_1+r)^2 (a_2+r^2)^{-2} + 2(a_1+r)(a_2+r^2)^{-1} \\
 f^{(2)}(R) &= 8r^2(a_1+r)^2 (a_2+r^2)^{-3} - 8r(a_1+r)(a_2+r^2)^{-2} - 2(a_1+r)^2 (a_2+r^2)^{-2} + 2(a_2+r^2)^{-1} \\
 f^{(3)}(R) &= -12(-a_2+2a_1r+r^2)(-a_1a_2 - 2a_2r+a_1r^2)(a_2+r^2)^{-4}
 \end{aligned}$$

$$\begin{aligned}
 gF(\nu_1, \nu_2) &= \nu_2(\nu_2 - 2)^{-1}, \nu_2 > 2 \\
 g(F - gF)^2 &= \text{Var}[F(\nu_1, \nu_2)] = 2\nu_2^2(\nu_1 + \nu_2 - 2)[\nu_1(\nu_2 - 2)^2(\nu_2 - 4)]^{-1}, \nu_2 > 4 \\
 g(F - gF)^3 &= 8\nu_1(\nu_1 + \nu_2 - 2)(2\nu_1 + \nu_1 - 2)(\nu_2 - 2)^{-3}(\nu_2 - 4)^{-1}(\nu_2 - 6)^{-1}, \nu_2 > 6
 \end{aligned}$$

## APPENDIX C

Compressed Template SAS IML<sup>®</sup> Program for Numerical Integration of Satterthwaite Power

```

PROC IML WORKSIZE=500 SYMSIZE=200;
  START F_OF_U(U,NU2,NU1,THETA); IF U<=0 THEN RETURN(0);
  NU1_2=NU1/2;      NU2_2=NU2/2;

```

```

T1=NU2 / (NU1*THETA) ;
L1=LGAMMA(NU1_2+NU2_2) - LGAMMA(NU1_2) - LGAMMA(NU2_2);
L2=NU2_2*LOG(T1); L3=(NU2_2-1)*LOG(U); L4=-(NU1_2+NU2_2)*LOG(1+T1*U);
LDEN=L1+L2+L3+L4;
IF LDEN<=-706 THEN RETURN(0); *Avoid underflow;
ELSE RETURN(EXP(LDEN)); FINISH F_OF_U;

START H_OF_U(U,N1,N2,NU1,NU2,THETA);
UP=(1/N1+U/N2) # (NU1+NU2);
DOWN=(NU1+NU2*U/THETA) # (1/N1+THETA/N2);
RETURN(UP/DOWN); FINISH H_OF_U;

START NU_OF_U(U,N1,N2,NU1,NU2);
T1=1/N1 + U/N2; T2=1/(N1*N1*NU1); T3=U*U/(N2*N2*NU2);
RETURN( T1*T1/(T2+T3) ); FINISH NU_OF_U;

START D_OF_Z(Z,ALPHA,N1,N2,THETA,LAMBDA);
NU1=N1-1; NU2=N2-1;
IF (Z<=0){Z>=1} THEN RETURN(0);
Z1=1/(1-Z); U=Z*Z1;
FCRITU=FINV(1-ALPHA,1,NU_OF_U(U,N1,N2,NU1,NU2));
HU=H_OF_U(U,N1,N2,NU1,NU2,THETA);
PU=1-PROBF(HU#FCRITU,1,NU1+NU2,LAMBDA);
DU=PU#F_OF_U(U,NU2,NU1,THETA);
RETURN( DU#Z1#Z1 ); FINISH D_OF_Z;

START SIMPSON5(A,B,FUNCTION,PARM_SET,NSTEP);
NSTEP2=2*NSTEP; DELTA=(B-A)/NSTEP2;
XSET = A + DELTA*(DO(0,NSTEP2,1)) ;
PARMSET=REPEAT(PARM_SET,NSTEP2+1,1);
T=APPLY(FUNCTION,XSET,PARMSET(,1),PARMSET(,2),PARMSET(,3),PARMSET(,4),PARMSET(,5));
SUM1=T(1,) + T(NSTEP2+1,);
I4=(DO(2,NSTEP2,2)); SUM4=SUM(T(I4,));
I2=(DO(3,NSTEP2+1,2)); SUM2=SUM(T(I2,));
RETURN( (SUM1 + 4*SUM4 + 2*SUM2)*DELTA/3 ); FINISH SIMPSON5;

STEP=30; ALPHA=.05; THETA=1; N1=6; N2= 6; LAMBDA=0;
PARM_SET= ALPHA||N1||N2||THETA||LAMBDA;
SATPOW=SIMPSON5((0),(1),"D_OF_Z",PARM_SET,NSTEP) ;
PRINT ALPHA THETA N1 N2 LAMBDA SATPOW ;

```

## BIBLIOGRAPHY

- Best DJ, Rayner JWC. Welch's Approximate Solution for the Behrens Fisher Problem. *Technometrics*. 1987;29, 205–211.
- Gans DJ. Use of a Preliminary Test in Comparing Two Sample Means. *Communications in Statistics—Simulation and Computation*. 1981;10, 163–174.
- Johnson, NL.; Kolz, S. *Continuous Univariate Distributions-2*. Boston: Houghton Mifflin; 1970.
- Johnson, NL.; Kotz, S. *Continuous Multivariate Distributions*. New York: Wiley; 1972.
- Lee AFS, Gurland J. Size and Power of Tests for Equality of Means of Two Normal Populations with Unequal Variances. *Journal of the American Statistical Association*. 1975;70, 933–941.
- Moser BK, Stevens GR. Homogeneity of Variance in the Two-Sample Means Test. *The American Statistician*. 1992;46, 19–21.
- Moser BK, Stevens GR, Watts CL. The Two-Sample T Test Versus Satterthwaite's Approximate F test. *Communications in Statistics—Theory and Methods*. 1989;18, 3963–3975.
- Muller KE, Barton CN. Approximate Power for Repeated-Measures ANOVA Lacking Sphericity. *Journal of the American Statistical Association*. 1989;84, 549–555.
- Muller KE, Barton CN. Correction to Approximate Power for Repeated-Measures ANOVA Lacking Sphericity. *Journal of the American Statistical Association*. 1991;86, 255–256.
- Zimmerman DW, Williams RH. Power Comparisons of the Student T-test and Two Approximations When Variances and Sample Sizes are Unequal. *Journal of the Indian Society of Agricultural Statistics*. 1989; XLI:206–217.

TABLE I

Differences Between Approximations and Exact Satterthwaite Power for  $\alpha = .05$

	N	Percentiles				
		Min	25	50	75	Max
Univariate Repeated	80	<b>-0.0088</b>	-0.0019	-0.0000	0.0006	0.0067
Naïve	150	-0.0094	0.0000	0.0006	0.0040	<b>0.0182</b>
Ratio of Expected Values	150	<b>-0.0179</b>	-0.0030	-0.0000	0.0001	0.0014
Taylor's Series, 1T	150	<b>-0.0158</b>	-0.0013	0.0001	0.0029	0.0116
Taylor's Series, 2T	150	-0.0018	0.0001	0.0025	0.0113	<b>0.0353</b>
Taylor's Series, 3T	104	-0.0018	0.0000	0.0016	0.0082	<b>0.0145</b>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE II

Exact and Approximate Error df Functions

$\theta$	$N_1$	$N_2$	$v$	$g_m$ v. Approximation					
				Naive	Exact	Ratio	TS1	TS2	TS3
	1	6	6	10.0	8.8	8.6	9.4	8.8	—
	1	6	51	6.2	7.2	5.9	6.2	6.8	6.8
	1	51	51	100.0	98.1	98.1	100.0	97.9	98.5
	10	6	6	6.0	6.4	5.7	5.6	7.5	—
	10	6	51	20.8	26.3	16.8	20.1	27.6	27.6
	10	51	6	5.1	5.2	5.1	5.1	5.2	5.2
	10	51	51	59.9	60.3	59.5	59.5	60.3	60.1